CLAIMS

What is claimed is:

1.   A method comprising:

receiving, into a capacity planning tool, configuration information for at least one streaming media server;

receiving, into said capacity planning tool, workload information for a workload of client accesses of streaming media files from a server; and

said capacity planning tool evaluating a capacity of the at least one streaming media server for supporting the workload.

2.   The method of claim 1 wherein said configuration information includes identification of size of memory of said at least one streaming media server.

3.   The method of claim 2 wherein said configuration information further includes disk configuration of said at least one streaming media server.

4.   The method of claim 1 wherein said workload information includes identification of number of concurrent client accesses of said streaming media files over a period of time.

5.   The method of claim 4 wherein said workload information further includes identification of a corresponding encoding bit rate of each of said streaming media files accessed.

6.   The method of claim 1 wherein said workload information comprises information from an access log collected over a period of time.

7.   The method of claim 1 wherein said evaluating comprises:

computing a cost corresponding to resources of said at least one streaming media server that are consumed in supporting the workload.

8.   The method of claim 7 wherein said computing said cost comprises:

computing a cost of consumed resources for a stream in said workload having a memory access to a streaming media file; and

computing a cost of consumed resources for a stream in said workload having a disk access to a streaming media file.

9.   The method of claim 1 wherein said evaluating comprises:

computing a service demand for said at least one streaming media server supporting said workload.

10.   The method of claim 9 wherein said computing said service demand comprises computing:

$$Demand = \sum_{i=1}^{K_W} N_{X_{W_i}}^{memory} \times cost_{X_{W_i}}^{memory} + \sum_{i=1}^{K_W} N_{X_{W_i}}^{disk} \times cost_{X_{W_i}}^{disk} \quad ,$$

wherein the workload W comprises $X_w = X_1,...,X_{k_W}$ set of different encoded bit rates of

files served in the workload, $N_{X_{W_i}}^{memory}$ is a number of streams in the workload having a memory

access to a subset of files encoded at X_{W_i} Kb/s, $cost_{X_{W_i}}^{memory}$ is a cost of consumed resources for a

stream having a memory access to a file encoded at X_{W_i} Kb/s, $N_{X_{W_i}}^{disk}$ is a number of streams in

the workload having a disk access to a subset of files encoded at X_{W_i} Kb/s, , and $cost_{X_{W_i}}^{disk}$ is a

cost of consumed resources for a stream having a disk access to a file encoded at X_{W_i} Kb/s .

11.   The method of claim 1 further comprising:

receiving at least one service parameter.

12.   The method of claim 11 wherein said at least one service parameter comprises information identifying at least one performance criteria desired to be satisfied by said at least one streaming media server under the workload.

13.   The method of claim 12 wherein said at least one performance criteria specifies a minimum percentage of time that said at least one streaming media server is desired to be capable of supporting the workload.

14.   The method of claim 11 wherein said at least one service parameter comprises information identifying a constraint.

15.   The method of claim 11 wherein said evaluating further comprises:

evaluating whether said at least one streaming media server satisfies said at least one service parameter.

16. Computer-executable software code stored to a computer-readable medium, the computer-executable software code comprising:

code for receiving workload information for a workload of client accesses of streaming media files from a server; and

code for evaluating a capacity of at least one system configuration for supporting the workload.

17. Computer-executable software code of claim 16 further comprising:

code for receiving configuration information for said at least one system configuration.

18. Computer-executable software code of claim 16 wherein said code for evaluating a capacity of at least one system configuration for supporting the workload comprises:

code for determining whether said at least one system configuration is capable of supporting said workload in accordance with at least one service parameter.

19. Computer-executable software code of claim 18 wherein said at least one service parameter comprises information identifying at least one performance criteria desired to be satisfied by said at least one system configuration under the workload.

20. Computer-executable software code of claim 16 further comprising:

code for generating a workload profile for the received workload information.

21. Computer-executable software code of claim 20 wherein the received workload information comprises an access log collected over a period of time.

22. Computer-executable software code of claim 20 wherein said workload profile comprises:

for a plurality of different points in time, identification of a number of concurrent client accesses, wherein the number of concurrent client accesses are categorized into corresponding encoding bit rates of streaming media files accessed thereby and are further sub-categorized into either memory or disk accesses.

23. Computer-executable software code of claim 16 wherein said code for evaluating comprises:

code for generating a service demand profile for said at least one system configuration.

24. Computer-executable software code of claim 16 wherein said code for evaluating a capacity of at least one system configuration comprises:

code for evaluating a capacity of a plurality of different system configurations and determining an optimal one of said plurality of different system configurations for supporting the workload.

25. A system comprising:

means for receiving configuration information for a plurality of different system configurations;

means for receiving workload information for a workload of client accesses of streaming media files from a server; and

means for evaluating the capacity of each of said plurality of different system configurations for supporting said workload.

26. The system of claim 25 further comprising:

means for determining an optimal one of said plurality of different system configurations for supporting said workload.

27. The system of claim 26 wherein said means for determining an optimal one of said plurality of different system configurations for supporting said workload determines a most cost-effective one of said plurality of different system configurations for supporting said workload according to determined service parameters.

28. A method comprising:

receiving workload information identifying an expected workload of client accesses of streaming media files from a server; and

determining a service demand profile for at least one server configuration under evaluation for evaluating a capacity of said at least one server configuration for supporting the expected workload.

29. The method of claim 28 further comprising:

receiving at least one service parameter.

30. The method of claim 29 wherein said at least one service parameter comprises information identifying at least one performance criteria desired to be satisfied by said at least one server configuration under the expected workload.

31. The method of claim 29 further comprising:

evaluating the determined service demand profile for the at least one server configuration to determine whether the at least one server configuration satisfies the received at least one service parameter.

32. A system comprising:

a media profiler operable to receive a client access log collected over a period of time for a service provider's site and generate a workload profile for the service provider's site; and

a capacity evaluator operable to receive the generated workload profile and evaluate at least one server configuration's capacity for supporting the site's workload.

33 The system of claim 32 wherein said capacity evaluator is further operable to receive configuration information for said at least one server configuration.

34. The system of claim 32 wherein in evaluating said at least one server configuration's capacity, said capacity evaluator determines whether said at least one server configuration is capable of supporting the site's workload in accordance with at least one service parameter.

35. The system of claim 34 wherein said at least one service parameter comprises information identifying at least one performance criteria desired to be satisfied by said at least one server configuration under the site's workload.

36. The system of claim 32 wherein said workload profile comprises:

for a plurality of different points in time, identification of a number of concurrent client accesses, wherein the number of concurrent client accesses are categorized into corresponding encoding bit rates of streaming media files accessed thereby and are further sub-categorized into either memory or disk accesses.

37. The system of claim 32 wherein in evaluating said at least one server configuration's capacity said capacity evaluator is operable to generate a service demand profile for said at least one server configuration.